# Shadow Segmentation and Shadow-Free Chromaticity via Markov Random Fields

*Cheng Lu and Mark S. Drew School of Computing Science, Simon Fraser University Vancouver, British Columbia, Canada*

## Abstract
*We design an algorithm based on illuminant invariance theory to find shadow regions in a colour image. Shadows are caused by a local change in both the colour and the intensity of illumination. Using both chromaticity and intensity cues, an illuminant discontinuity measure is derived by which shadow edges can be locally identified. We model the problem of finding shadows by a Markov Random Field using our new measure. A graph-cut optimization method is then applied to the MRF to find the globally optimal segmentation of shadows in an image. In previous work, a 2-d chromaticity colour invariant image was recovered from a greyscale 1-d invariant image by adding back light so as to match the chromaticity of bright pixels. Here, since we segment shadows, we can take a completely different approach and leave nonshadow pixels unchanged, while adding light to shadow pixels so as to match neighbouring nonshadow pixels. The results are much more convincing shadow-free images, and shadow-segmentation is excellent.*

## 1. Introduction
Many computer vision algorithms, such as segmentation, tracking, and stereo registration, are confounded by shadows in images. Hence finding or removing shadows in colour images is an important research issue. The problem is difficult due to the confounding interaction between object reflectance and illumination. An interesting feature of this problem is that shadows are approximately but accurately described as a change of lighting. Hence, it is possible to cast the problem of finding shadows in images into an equivalent statement about detecting changes in lighting in imagery.

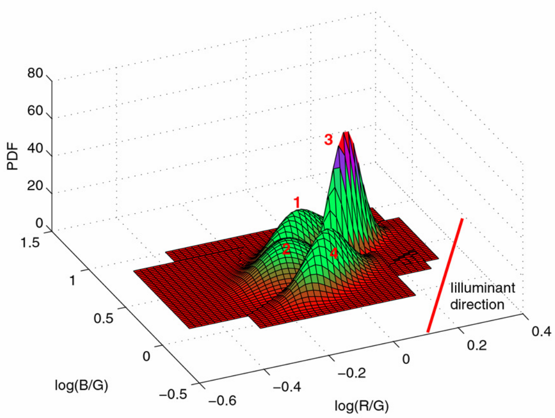
A recent method was devised for creating an illumination-invariant image from an input colour image,[1–3] invariant to both colour and intensity of the scene lighting. The method is in essence a kind of calibration scheme for a particular colour camera. A camera is calibrated by imaging a target which contains a set of colour patches, under several different illuminants. An invariant image is derived based on the idea that under Planckian lighting, and for camera sensors that are narrowband (as for an ideal delta-function sensor camera) a 2-d scatter plot of the logarithms of ratios $R/G$ versus $B/G$, say, produce a set of approximately straight lines (this is the case for any model of illumination that changes light colour by exponentiation of a power of temperature, $T$). Each line corresponds to a single patch of the target; each point on a line corresponds to a particular illuminant. For a given camera, all such lines are essentially parallel. We call the direction of these straight lines the *illuminant direction*. (Note that gamma-correction does not change the straight line theory.[1]) The invariant image is formed by projecting 2-vectors onto the direction orthogonal to this direction.

In Fig. 1(a), the photographer's cast shadow lies on the path and the grass. Fig. 1(b) plots the probability densities of the 2-d log ratio chromaticity points, and the illuminant direction, which is parallel to the red line. There are four main regions in this image: lit path, shadow path, lit grass and shadow grass, marked 1-4 in Fig. 1(a), corresponding to the four main concentration peaks in Fig. 1(b). When projecting chromaticity points into the direction orthogonal to the red line, the four clusters fall into two portions: one for lit grass and shadow grass; and the other for lit path and shadow path. Thus, points in shadow have the same value as points in corresponding nonshadow regions. Fig. 1(c) shows the result of this projection as a 1-d greyscale image, with shadows removed very well.

Note that the projection operation eliminates not only shadow effects but also all information along the illuminant direction. Many factors affect the distribution of the chromaticity along the illuminant direction, such as noise and surface texture; as well, real cameras have non narrowband sensors — these factors can make the distribution complex. A straightforward observation is that in Fig. 1(b), the four regions have many overlapping points, although the majority are separated. It is expected that when points are projected into 1-d quantities, much information which is irrelevant to the illuminant will coalesce. This will prove troublesome for detecting edges in the invariant image, which in turn may impact the accuracy of shadow edge detection, and hence make localizing shadow regions difficult.[4]

Figure 1(b) shows that the four clusters constitute two pairs, each of which aligns along the illuminant direction: a shadow-nonshadow grass pair, and a shadow-nonshadow path pair. This leads to a general observation: rather than projecting all 2-d chromaticities into 1-d quantities, we can simply compare pixels on either side of a region edge: if their chromaticities belong to one of the pairs, the edge is probably a shadow edge. We call two pixels with a neighbour relation across a shadow-nonshadow edge an illuminant discontinuity pair. In this paper we propose a continuous angle between vectors as a criterion for measuring the illuminant discontinuity. We then define a Markov Random Field to model our binary shadow segmentation problem. The MRF approach combines spatial context and local chromaticity features to identify shadow-nonshadow edges. The segmented shadows are found via energy minimization of this random field using a Graph Cut algorithm.[5]


(a)




(b)

(c)

*Figure 1. (a) Original colour image. The four main regions 1-4 are marked. (b) Log band-ratio chromaticities probability densities. (c) Greyscale invariant image.*

Previous work[6] extends the illuminant invariance method from 1-d greyscale output to output which is 2-dimensional colour, in the form of $L_1$ normalized chromaticity $\chi$, defined as[7]:

$$c = \{r, g, b\} \equiv \{R, G, B\}/(R + G + B) \tag{1}$$

For a Lambertian surface the 2-d chromaticity removes shading and intensity from images, but still conveys colour information. Although not a full-colour result, as in Ref. [3], the $L_1$ normalized chromaticity space has still been widely used to reduce shading and shadow effects in some vision applications, since chromaticity space can indeed remove shadows due to changes in lighting intensity only. However, in case of shadows being caused by changes of lighting colour, shadows will not be eliminated (see Fig. 4(a)).

To obtain an invariant $L_1$ normalized chromaticity image, the method in Ref. [6] consists of preserving the 2-vector components for the 1-d projected log ratio chromaticities, followed by restoring lighting to the image so as to best match the brightest pixels of the original image. However, the resulting 2-vectors still contain only 1-d information—we simply shift all chromaticities on the projection line by a fixed amount of chromaticity, hoping to restore light in the shadows while at the same time matching the chromaticity for bright pixels.

Here, we take another tack entirely: We first segment the shadow/nonshadow image using the proposed MRF. Then since shadow regions are indeed found we can now restore the lighting to pixels in shadows but leave the chromaticity outside the shadow unchanged. This should produce a much more convincing chromaticity image for the invariant image since we leave untouched information we do not in fact wish to alter. Our approach is to evaluate an offset for a shadow pixel by calculating the distance between shadow and nonshadow points in the log ratio space, and then apply the offset to shadow points so as to move their log ratios to the corresponding nonshadow ones. Thus, the resulting chromaticity image resists shadow effects but leaves nonshadow regions unchanged.

## 2. Illuminant Invariant Formation

We recapitulate how linear behaviour with lighting change results from the assumptions of Planckian lighting, Lambertian surfaces, and a narrowband camera. Consider the RGB colour 3-vector $\rho$ formed at a pixel with illumination with spectral power distribution $E(\lambda)$ impinging on a surface with surface spectral reflectance function $S(\lambda)$. If the three camera sensor sensitivity functions form a set $\boldsymbol{Q}(\lambda)$, then we have

$$\rho_k = \sigma \int E(\lambda)S(\lambda)Q_k(\lambda)d\lambda,\ k = R, G, B, \tag{2}$$

where $\sigma$ is Lambertian shading: surface normal dotted into illumination direction.

If the camera sensor $Q_k(\lambda)$ is exactly a Dirac delta function $Q_k(\lambda) = q_k\delta(\lambda - \lambda_k)$, then Eq. (2) becomes simply

$$\rho_k = \sigma E(\lambda_k)S(\lambda_k)q_k,\ k = 1 \ldots 3. \tag{3}$$

Now suppose lighting can be approximated by Planck's law, in Wien's approximation[7]

$$E(\lambda,T) \simeq I\, c_1 \lambda^{-5} e^{-\frac{c_2}{T\lambda}}, \tag{4}$$

with constants $c_1$ and $c_2$. Temperature $T$ characterizes the lighting colour and $I$ gives the overall light intensity.

In this approximation, from Eq. (3) the RGB colour $\rho_k$, $k = 1...3$, is simply given by

$$\rho_k = \sigma I c_1 \lambda_k^{-5} e^{-\frac{c_2}{T\lambda_k}} S(\lambda_k) q_k . \tag{5}$$

Let us now form the band-ratio 2-vectors $\boldsymbol{r}$,

$$r_\mu = \rho_\mu / \rho_p, \tag{6}$$

where $p$ is one of the channels and $\mu$ indexes over the remaining responses. We could use $p = 2$ (i.e., divide by Green) and so calculate $r_1 = R/G$ and $r_2 = B/G$. We see from Eq. (5) that forming the chromaticity effectively removes intensity and shading information. If we now form the log of Eq. (6),

$$r'_\mu \equiv \log(r_\mu) = \log(s_\mu / s_p) + (e_\mu - e_p)/T, \ \mu = 1,2 \tag{7}$$

with $s_\mu \equiv c_1 \lambda_\mu^{-5} S(\lambda_\mu) q_\mu$ and $e_\mu \equiv -c_2/\lambda_\mu$. Thus Eq. (7) is a straight line parameterized by $T$. Notice that the 2-vector direction $(e_\mu - e_p)$ is *independent of the surface* — it captures the illumination direction.

The invariant image is that formed by projecting 2-d logs of chromaticity, $r'_\mu, \mu = 1,2$, into the invariant direction $e^\perp$ orthogonal to the vector $e \equiv (e_\mu - e_p)$. The result of this projection is a single scalar which we then code as a greyscale value.

The utility of the invariant image is that it reveals how the changes of lighting cause shadow effects, and provides a way to factor the interaction between the lighting and surface in the log ratio chromaticity space, such that lighting changes can be removed in the resulting invariant image.

## 3. Illuminant Discontinuity Measure

From the above section, we see that two pixels of a single surface under two different lights will align with the direction $(e_\mu - e_p)$ in the log ratio chromaticity space. In the image space, on the other hand, they will appear with different intensities, and thus the shadow effect occurs. If these two pixels are close to each other in the image, they are an *illuminant discontinuity pair*. Now, we are aiming at specifying a criterion by which illuminant discontinuity pairs can be determined.

For a particular camera, the illuminant direction can be computed either by a pre-calibration scheme[3] or an automatic entropy minimization method.[4] We denote the illuminant direction 2-vector by $e_0$. Suppose a pair of pixels $\{i,j\}$ are neighbouring in the image and have quite different intensities; we now try to measure how close to $e_0$ is the direction of the vector linking the two pixels in the log feature space. We give a high score to a linking vector with direction similar to $e_0$, and so propose an *illuminant discontinuity measure*

$$\kappa_{ij} = \left| \frac{< e_{ij}, e_0 >}{\left\| e_{ij} \right\| \left\| e_0 \right\|} \right| \tag{8}$$

where $e_{ij}$ is the vector linking the two log ratios of neighbouring pixels $i,j$. We take the absolute value of the cos because we consider only the angles between 0 and $\pi/2$.

The utility of the discontinuity measure is that it is continuous, rather than a discrete {admit/reject} variable, and will prove beneficial in the later graph cut optimization for finding shadows.

## 4. Finding Shadows Using MRF and Graph Cut Optimization

Now we try to assign labels (shadow and nonshadow) to pixels based on the illuminant discontinuity measure. In the label processing, the labels tend to vary smoothly within the image except at region boundaries, where discontinuities occur. The fact that a particular pixel label depends on the labels of its neighbours allows modeling the optimization problem as a Markov Random Field. The MRF-based segmentation model is defined by the contextual relationships within a local neighbourhood structure. Since our goal is the assertion of local discontinuity constraints, we will consider only first order random fields, both simplifying the model and limiting the computational complexity.

The formulation of our MRF model will be similar to others used for segmentation except for a number of variations due to the characteristics of our illuminant discontinuity measure. Suppose we are given a colour image pixel set $X = \{i\}$, on a first order neighbourhood system represented by a set $\mathcal{N}$ of all unordered pairs, and each site $i$ takes a label $l_i$ in $L = \{0,1\}$, with $\{0,1\}$ denoting labels "shadow" and "nonshadow". Then we use a Potts cost function to describe the MRF model:

$$E(l) = \sum_{i \in X} D_i(l_i) + \sum_{\{i,j\} \in \mathcal{N}} B_{ij} \cdot \delta(l_i, l_j) , \tag{9}$$

and

$$\delta(l_i, l_j) = \begin{cases} 1 & \text{if } l_i \neq l_j \\ 0 & \text{otherwise} \end{cases}$$

The data term $D_i$ states that the costs for assigning pixel ito "shadow" and "nonshadow" are respectively $D_i(0)$ and $D_i(1)$. The term $B_{ij}$ reflects the discontinuity property between $i$ and $j$. Normally, $B_{ij}$ is large when pixels $i$ and $j$ are similar, but $B_{ij}$ is close to zero when a discontinuity occurs between them. Finding the best labeling forms an optimization problem. The binary labeling problem for the Potts energy function as formulated above can be solved, optimally, by a single execution of the Graph Cut algorithm[5] to find the max-flow solution. For the purpose of using graph cuts, we create a graph with nodes corresponding to pixels. There are two additional terminal nodes: a "shadow" terminal (a source) and a "nonshadow" terminal (a sink). All the links in the neighbourhood system $\mathcal{N}$ are edges in the graph. These edges are called *n-links*. Edges between pixel nodes and the terminal nodes are added to the graph and called *t-links*. These t-links are assigned costs based on the data term $D_i$, while n-links are assigned costs based on the boundary term $B_{ij}$.

In our shadow labeling problem, there is no cue to indicate the likelihood of a shadow or nonshadow label for each individual pixel, so we assign the same cost to each t-link, thus actually giving up the data term in the model, for t-links. The data term thus imposes no influence on the later minimum-cut calculation. Now the task is to find the shadow-nonshadow graph cut and a label for each pixel node, by setting the cost for n-links. For this purpose, we use directed n-links ($i, j$)and ($j, i$) between neighbouring pixels $i$ and $j$ (see Fig. 2).

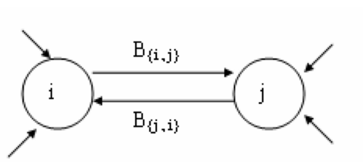



*Figure 2. Two nodes with directed n-links.*

The cost function $B_{ij}$ can be defined as:

$$B_{ij} = K - \kappa_{ij}(I_j - I_i) \quad (10)$$

and

$$K = \max_{\{i,j\}\in\mathcal{N}} \kappa_{ij}|I_i - I_j|$$

where $I_i$ and $I_j$ are intensities at pixels $i$ and $j$ so as to incorporate both gradient direction and gradient magnitude. A minimum cut is a subset of edges that separates the source "shadow" from the sink "nonshadow" where the sum of its edge costs is minimum. This cost function $B_{ij} > 0$should be interpreted as penalty for a cut between $i$ and $j$. From Eq. (10), $B_{ij}$ gives a high penalty to a cut between pixels which have small illuminant discontinuity measure or similar intensities. Neighbouring pixels with high positive intensity gradient and high illuminant discontinuity correspond to low-cost n-links, which are attractive choices for the optimal segmentation boundary. However, the penalty $B_{ij}$ is much increased when the intensity gradient is negative, i.e. $I_j - I_i < 0$: this forces the "shadow" and "nonshadow" labeling for neighbouring pixels to be consistent with their intensity gradient direction — dark pixels stay in shadow and bright pixels stay in nonshadow segments. Figure 3 shows shadows found for the image in Fig. 1(a). Results are excellent.

*Figure 3. Shadows found using graph cut.*

The model Eq. (10) gives us a coherent representation of the labeling problem by incorporating both illuminant colour and intensity constraints. However, it omits one important point: not all shadow-nonshadow pairs have low cost. As we have mentioned, shadows result from changes in colour and intensity of the scene lighting. If, in some conditions, intensity changes dominate shadow generation, i.e. the contribution of illuminant colour change to the shadow effect is very small, the log ratio chromaticity space will of itself be able to greatly attenuate shadows, leading to shadow-nonshadow pairs with similar chromaticities. Two very close chromaticities will result in a nearly random measure $\kappa$, which in turn make the link cost untrustworthy. To overcome this drawback, we introduce a threshold $\tau$ for the distance between two points in the log ratio chromaticity space, such that if their distance is less than $\tau$, we do not evaluate the discontinuity measure.

But we still have to assign a value to the $\kappa$ for setting the n-link cost. We have two choices: $\kappa_{i,j} = 1$ and $\kappa_{i,j} = 0$, for two neighbouring pixels $i$ and $j$ which are close in chromaticity space. The former setting — for two neighbouring pixels with similar chromaticities but large intensity gradient — will be linked by a low cost edge such that a cut is prone to happen there. This setting is based on the fact that if a region has a lower intensity than its neighbouring region, but has the same chromaticity, the region is probably in shadow. On the other hand, if we choose the latter setting, a pure intensity change will not be considered as due to a shadow and the cost function will put a high penalty on a cut happening on the link. For images where the shadow is completely caused by intensity changes we segment the shadow by choosing the former setting.

The minimum cut can be computed exactly in polynomial time using well known algorithms for 2-terminal graph cuts, e.g. max-flow.[8] The primary drawback is that texture and noise may confuse the discontinuity measure. To remove these factors, the Mean Shift method is used to filter the image first, and edges are then detected on the filtered image. Our illuminant discontinuity criterion and intensity gradient is calculated on only these edges, yielding a fast n-links cost computing. A second undesired effect is that shadow edges usually are not sharp but diffuse, so that the chromaticity of neighbouring pixels may be too close. To overcome this effect, for two pixels across an edge the illuminant discontinuity criterion is calculated using the means of blocks of pixels on the two sides of the edge.

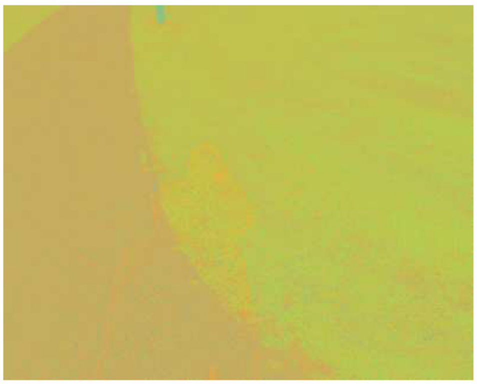
## 5. Shadow-Free Chromaticity Image

Drew et al.[6] formed an invariant $L_1$ normalized chromaticity ($\chi$ in Eq. (1) for a colour image based on the projected 1-d log ratio chromaticities. This is 2-d colour since

$$\sum_{k=1}^{3} \chi_k = 1.$$

Figure 4(a) shows the $\chi$ image: the shadow is clearly visible. Here we propose a new scheme to build the chromaticity image which puts lighting back into the shadow pixels but preserves the original nonshadow chromaticity.

(a)



(b)

*Figure 4. (a). Original chromaticity image. (b). Shadow-free chromaticity image.*

We make a simple yet in practice important observation. Consider a surface under two different lights, with temperatures Ta and Tb respectively. From Eq. (7), we have two log ratio chromaticity vectors:

$$r'_{a\mu} = \log(r_{a\mu}) = \log(s_\mu / s_p) + (e_\mu - e_p)/T_a,$$
$$r'_{b\mu} = \log(r_{b\mu}) = \log(s_\mu / s_p) + (e_\mu - e_p)/T_b, \tag{11}$$

with $\mu = 1, 2$. Subtracting,

$$r'_{a\mu} - r'_{b\mu} = (e_\mu - e_p)(1/T_a - 1/T_b), \tag{12}$$

so, clearly, the distance between the two chromaticity vectors depends on only the illuminant change. Thus we can estimate a single offset vector based on the distance between the shadow and nonshadow points in the log ratio chromaticity space, and then put back the offset to shadow points so as to move their log ratio chromaticities to the nonshadow ones. To do this, we collect a number of pixels inside and outside the shadow boundary, calculate the mean vectors for the shadow points and the nonshadow points respectively, and then subtract the local shadow mean vector from the nonshadow one to obtain the offset vector $r'_{offset}$. We now can add the offset vector locally to the shadow chromaticities, and then go to exponentiated values:

$$r'(shadows) = r'(shadows) + r'_{offset},$$
$$r = \exp(r'). \tag{13}$$

The result is band-ratio chromaticity. To obtain $L_1$ normalized chromaticity $\chi$, we have from Eq. (1) that

$$\chi = \{R, G, B\}/(R, G, B)$$
$$= \{r_1, 1, r_2\}/(r_1 + 1 + r_2), \tag{14}$$

so that in fact knowing r gives us $\chi$ as well. Fig. 4(b) shows the effect of restoration of the 2-vector offset required to put the nonshadow lighting back into the recovered chromaticity. We see that the method indeed does very well, compared to the shadowed, original version Fig. 4(a).

## 6. Additional Experiments

Figures 5 and 6 show more results. Using MRF and graph cut techniques, shadows can be extracted fast, and accurately. Adding back light into just the shadow regions produces an excellent shadow-free chromaticity image.

## 7. Summary

We have presented a Markov Random Field-based model for finding shadows. This model uses an illuminant discontinuity measure criterion along with intensities for labeling shadows using graph cut optimization. Based on the shadow regions found, we form an $L_1$ normalized chromaticity image which greatly attenuates shadow effects while keeping nonshadow pixels unchanged. Results are seen to be excellent. However, the images we tried all have shadow regions good for the task at hand: large area, sharp boundaries, and relatively clean background. One problem not addressed so far is how well the method performs in more complicated scenes. Also, we evaluate the illuminant discontinuity measure based on prior knowledge of the illuminant direction, which may not be accurate. In future work, we shall use the shadows found to adjust the illuminant direction, and iterate to produce an optimal solution for both shadows and lighting.

## References


1. G.D.Finlayson and S.D.Hordley. Color constancy at a pixel. J. Opt. Soc. Am. A, 18(2):253–264, 2001.
2. G.D. Finlayson and M.S. Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting and specularities. In ICCV, pages II:473–480. IEEE, 2001.
3. G.D. Finlayson, S.D. Hordley, and M.S. Drew. Removing shadows from images. In ECCV, pages 4:823– 836, 2002.
4. G.D. Finlayson, M.S. Drew, and Cheng Lu. Intrinsic images by entropy minimization. In ECCV, pages 582–595, 2004.
5. Y. Boykov, O. Veksker, and R. Zabih. Markov random fields with efficient approximation. In CVPR, pages 648–655, 1998.
6. M.S. Drew, G.D. Finlayson, and S.D. Hordley. Recovery of chromaticity image free from shadows via illumination invariance. In IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV'03, pages 32–39, 2003.
7. G.Wyszecki and W.S. Stiles. Concepts and Methods, Quantitative Data and Formulas. Wiley, New York, 2nd ed., 1982.
8. L. Ford and D. Fulkerson. Flows in Networks. Princeton Univ. Press, 1962.